

WORKSHEET

MACHINE LEARNING ANSWERSHEET 5

1. R-Squared value is used to measure the goodness of fit or best fit line. The greater the value of R-squared, the better is the regression model as most of the variation of actual values from the mean value get explained by the regression model. The residual sum of squares(RSS) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation.

2. TSS, ESS and RSS:

- ◆ TSS is total sum of squares, ESS is explained sum of squares and RSS is residual sum of squares.
- ◆ TSS measures the variation in values of an observed variable with respect to its sample MEAN, ESS calculates the deviation between the estimated value and the MEAN value of the observed variable. RSS is equal to sum of squares of all the errors or residual made by the model on the data.
- ◆ $TSS = ESS + RSS$

- 3. It minimizes the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.**
- 4. The Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset. Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.**
- 5. Decision trees are extremely prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. Overfitting occurs when the learning algorithm continues to develop suggestions that reduce training set error at the cost of increased test set error.**

- 6. Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has improved the popularity of ensemble methods in machine learning.**
- 7. Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types. Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.**
- 8. The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained.**

9. **K-fold cross-validation, you split the input data into k subsets of data. You train an ML model on all but one ($k-1$) of the subsets, and then evaluate the model on the subset that was not used for training. when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data is known as K-fold cross validation.**
10. **Hyperparameter tuning consists of finding a set of finest hyperparameter values for a learning algorithm while applying this improved algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.**
11. **If learning rate that is too large, then gradient descent algorithm can diverge from the optimal solution. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution. The gradient may simply keep oscillating around the optimal solution if the learning rate is high and it will not settle at the optimal solution.**

- 12. If we have Non-Linear Data where we have non-Linear decision boundaries, Then if we try to use the logistic regression it will perform poor on the data as the decision boundary in the data is nonlinear. Therefore, we cannot use Logistics regression for classification of Non-Linear Data because the decision boundary produced by Logistics Regression is Linear.**
- 13. AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.**
- 14. The bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters. This tradeoff happens between error made by the model and how much the model changes with changes in training data when model complexity changes. If model does not have much Bias neither much high variance and it carried out only at a point where we tradeoff between bias and variance.**

15. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example, linear, polynomial, radial basis function (RBF).

- ◆ Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line.**
- ◆ Polynomial Kernel is used when the data is the form of the polynomial of some degree n . is often used in SVM classification problems where the data is not linearly separable.**
- ◆ RBF is used when the data follows some complex patterns which is neither linear nor polynomial.**